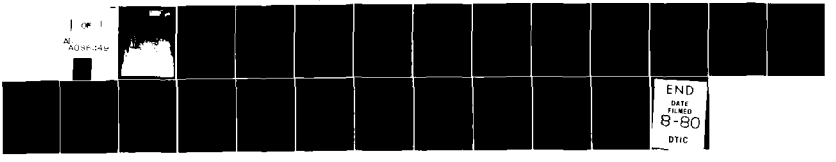
AD-A086 449 DESMATICS INC STATE COLLEGE PA F/G 12/2
FACTOR SCREENING IN COMPUTER SIMULATION: CONSIDERATIONS IN PERF--ETC(U)
JUL 80 D E SMITH, C A MAURO N00014-79-C-0650
UNCLASSIFIED TR-113-2 NL



DESMATICS, INC.

P. O. Box 618
State College, Pa. 16801
Phone: (814) 238-9621

Applied Research in Statistics - Mathematics - Operations Research

FACTOR SCREENING IN COMPUTER SIMULATION: CONSIDERATIONS IN PERFORMANCE EVALUATION

by

Dennis E. Smith

Carl A. Mauro

TECHNICAL REPORT NO. 113-2

July 1980

This study was supported by the Office of Naval Research under Contract No. N00014-79-C-0650, Task No. NR 274-317

TABLE OF CONTENTS

ABSTRACT

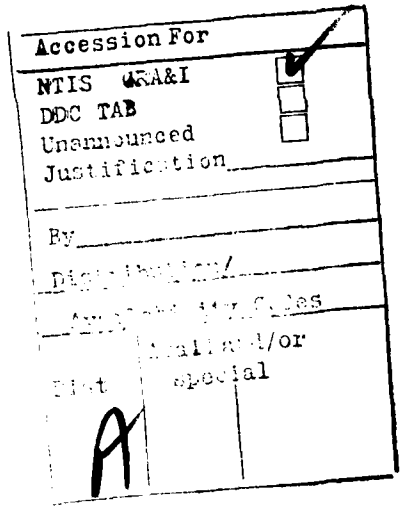
Although computer simulation is a major area in which factor screening situations are frequently encountered, adequate methodology has not been developed to resolve the factor screening problem. Within the constraint of a limited number of computer runs, a decision must be made about the selection of a screening strategy that will perform efficiently and effectively. This report reviews the major classes of screening designs that have been suggested and recommends two strategies for further research. In addition, a fundamental screening model is defined, and a performance measure for evaluating screening strategies is developed.

Key Words:

Factor Screening

Computer Simulation

Experimental Design



I. INTRODUCTION

At the beginning of an experimental investigation there are often a large number of factors (i.e., independent variables) to be analyzed, but only a limited number of experimental runs available. In most experimental situations of this type, it is anticipated that only a small proportion of the many original factors are important. Accordingly, the usual purpose of these experiments is to screen the factors in hope of identifying the relatively few that do have an appreciable effect on the response (i.e., dependent variable). Experiments having this intent are known as factor screening experiments and can arise in practically any field of scientific research. Screening experiments are often performed in the early stages in order to help identify potentially important factors to be subjected to more intensive investigation in subsequent experimentation.

Under the constraint of a limited number of experimental runs, it is difficult (if not impossible) to thoroughly investigate all the factors originally under consideration. Consequently, the design and analysis of screening experiments can present special problems that require an extension of conventional experimental design theory. Although factor screening may be based on undesigned data, the framework of interest in this report is the designed experiment case since it allows factor screening to be based on "planned" data and thus permits a more efficient statistical assessment of factor effects.

Computer simulation is a major area in which factor screening situations

are encountered. It is not unusual, for example, for a simulation user to wish to efficiently and effectively screen 100 factors using only 40 computer runs. Nonetheless, there has been only sparing research on factor screening in the computer simulation environment. To date, no specific methodology has been developed to adequately resolve the factor screening problem, although a number of general suggestions have been made [e.g., Hunter and Naylor (1970), Jacoby and Harrison (1962), Kleijnen (1975a), Smith (1973), and Montgomery (1979)]. Furthermore, no data comparing the performance of different screening strategies currently exists. This report recommends two strategies for detailed investigation and suggests a performance measure to be used for evaluating them or any other strategy.

In devising factor screening strategies for use in computer simulation experiments, it is desirable to provide a common mathematical basis for any screening strategy that might be proposed. Thus, the following paragraphs summarize the fundamental factor screening model assumed to underlie the simulation responses.

Suppose that each of the K factors to be screened are at two levels, arbitrarily designated "high" and "low." Define as active any factor which produces a nonzero change in the mean response. For analysis purposes, assume (without loss of generality) that the first k indexed factors are active and the remaining K - k factors are inactive.

Define

$$x_{ij} = \begin{cases} +1, \text{ if factor } j \text{ is set at its high level for the } i^{\underline{th}} \\ \text{computer run} \\ -1, \text{ if factor } j \text{ is set at its low level for the } i^{\underline{th}} \\ \text{computer run} \end{cases}$$

and let y_i equal the response for the $i^{\underline{th}}$ computer run. The factor screening model assumes that

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i,$$

where β_j is the linear effect of factor j and the error terms, ε_i, are independent, identically distributed $N(0, \sigma^2)$ random variables.

It should be noted that:

(a) It is implicit in the model that $\beta_j = 0$ when $j > k$. That is, if factor j is inactive, then $\beta_j = 0$.

(b) Under the parameterization of the model, the main effect of factor j equals $2\beta_j$. (The main effect of a factor is the difference between the average response at its high level and the average response at its low level.)

(c) The model is additive; that is, no interactions are assumed.

(d) In most cases, the number of computer runs available for screening is <u>less</u> than K, the number of factors to be screened.

II. OVERVIEW OF SCREENING DESIGNS

Two major classes of screening designs have been suggested in the literature. [See Kleijnen (1975b) for a summary.] These are: (1) the Plackett-Burman designs and (2) the supersaturated designs. The Plackett-Burman designs, which require more computer runs than factors to be screened, provide unconfounded estimates of the factor effects. The supersaturated designs, on the other hand, require no minimum number of runs, but do result in confounded estimates.

In the vast majority of simulation studies, the number of computer runs available for screening is considerably less than the number of factors to be screened. Thus, the Plackett-Burman designs are not well-suited, per se, to the factor screening problem. However, they are extremely important because they can serve as a complementary component to supersaturated designs in various screening strategies. Both classes of designs are briefly reviewed in this section.

A. PLACKETT-BURMAN DESIGNS

Plackett-Burman designs [Plackett and Burman (1946)] are two-level orthogonal, fractional factorial designs for studying up to $K = 4m - 1$ factors in $N = 4m$ runs. Because of the orthogonality, no main effect is confounded with any other main effect. Hence, if all interactions are negligible (as in the fundamental factor screening model), unbiased estimation of the K factor effects is possible.

Furthermore, with respect to orthogonality, Plackett-Burman designs

are minimal in the necessary number of runs. Indeed, the orthogonality condition could not be satisfied if $N < K + 1$, for otherwise the columns of the design matrix would need to form a set of $K + 1$ orthogonal vectors in an $N < K + 1$ dimensional space. Because of these properties, Plackett-Burman designs would undoubtedly be the design of choice if $N > K$ runs could be afforded. Unfortunately, factor screening problems usually do not provide this luxury.

B. SUPERSATURATED DESIGNS

It is well known that as a design property, orthogonality is a desirable feature since it allows for independent estimation of factor effects. Furthermore, if two columns are orthogonal, the corresponding factors are unconfounded. That is, if one of the factors has an effect then it cannot cause the other factor to appear (perhaps erroneously) to have an effect. As already noted, in order that all columns be mutually orthogonal, N must exceed K. When $N \leq K$, the design is called supersaturated. There are three main classes of supersaturated designs which have been suggested for factor screening. These are reviewed in the following sections.

1. Random Balance Designs

A random design is one for which a random sampling process is used to choose all or some of the elements of the design matrix. Such designs are described by Anscombe (1959), Budne (1959), and Satterthwaite (1959). In order to generate the design matrix, various sampling schemes may be employed. The most appealing procedure for a two-level situation is to consider only

balanced designs in which the high and the low levels of a factor appear equally often. That is, every column of the design matrix (except the initial column of 1's) consists of N/2 +1's and N/2 -1's (N even). This guarantees that the factor effect estimates are unconfounded with the overall mean effect. The actual assignment of +1's and -1's is made randomly, with each column receiving an independent assignment. Note that as opposed to Plackett-Burman designs, N can be determined independently of K. This is a desirable feature for any screening design to have.

At the basis of objection to random designs is the random degree of confounding that such designs introduce. Furthermore, there is no specific method of analysis for these designs. A complete discussion of the merits and demerits of random designs can be found in Budne (1959), Satterthwaite (1959), and Youden, Kempthorne, Tukey, Box, and Hunter (1959).

2. Systematic Supersaturated Designs

Because of the random degree of confounding that occurs in random balance designs, Booth and Cox (1962) introduced systematic two-level factorial designs which, in some sense, minimize confounding. They considered designs where, as for a random balance design, each of the K factors appear at the low and the high levels an equal number of times. Letting $\underline{x}_i$ denote the design column of factor i, recall that design columns i and j are orthogonal if and only if $\underline{x}_i'\underline{x}_j = 0$. Since $N \leq K$, not all the design columns can be orthogonal. With this in mind, Booth and Cox constructed designs that minimized $\max_{i \neq j} |\underline{x}_i'\underline{x}_j|$, and they tabulated the resulting designs for various values of N and K, where $K \leq 36$.

For larger K, an iterative computer procedure is outlined for generating the required designs. However, the cost required to write and to run such a program can be large. Furthermore, as Booth and Cox indicate, systematic supersaturated designs seem to provide little advantage over random designs when K is larger than 2N. Hence, for screening situations in simulation studies where this is often the case, random balance designs appear more attractive than systematic supersaturated designs.

3. Group Screening Designs

The basic idea underlying group screening designs is to partition the factors into groups and then to consider each group of factors as a single factor. The resulting smaller number of group-factors can then be tested efficiently in a conventional design such as a Plackett-Burman design. As originally proposed by Watson (1961), group screening involves a two-stage process, although Patel (1962) and Li (1962) have generalized group screening to more than two stages.

In group screening, group-factor levels are defined by equating the levels of component factors to the group-factor level. Hence, the grouping procedure induces complete confounding among the factors within a group. Accordingly, in the two-stage procedure, only those individual factors comprising significant group-factors are examined in the second stage. In multistage group screening, each succeeding stage consits of repartitioned groups of the significant group-factors from the previous stage.

In the development of group screening procedures, a number of assumptions were made. A primary one is that the directions of possible effects are

known, a priori. This assumption assures against possible cancellation of a group-factor effect. However, whether or not group screening can be used effectively even when the direction of the potential effects are not known is an important issue that will be addressed in a forthcoming Desmatics report.

III. TWO PROPOSED STRATEGIES

As noted in Section I, no systematic evaluation of factor screening methods presently exists. Nevertheless, there have been some attempts to establish guidelines for the choice of a factor screening strategy. When K is large and N is of moderate size, Kleijnen (1975a) and Montgomery (1979) suggest that group screening might be the more useful procedure. However, their recommendations are based mostly on subjective assessments. Their apparent lack of support for random balance designs appears to stem from the uncertainty in the effect of random confounding on the performance of screening strategies associated with random designs as well as the lack of specific methodology with which to analyze random balance data. Although much of this insecurity may be real, there is a lack of analytical and empirical results to adequately resolve the issue. Further research is needed.

After consideration of the various designs reviewed in the previous section, two main screening strategies were identified as warranting more rigorous investigation. The first is based on the two-stage group screening design, while the second is a two-stage combination strategy based on random balance and Plackett-Burman designs. In order to maintain flexibility, the total number of runs, R, necessary for each strategy is allowed to be a random variable. The two basic strategies are outlined in the following sections.

A. TWO-STAGE GROUP SCREENING STRATEGY

The initial step in this strategy is to randomly partition the K factors

into groups of size g. If K is not a multiple of g, the group sizes are chosen as "evenly" as possible. For example, if K = 46 and g = 3, the partition would consist of 14 groups of size 3 and one group of size 4 for a total of 15 groups. Let G denote the number of groups into which the K factors are partitioned.

Once the factors have been partitioned into G groups, the group-factors are systematically tested using a Plackett-Burman design employing N runs. The number of first stage runs, N, will equal the smallest integer larger than G which is a multiple of four. For instance, in the above example, N = 16.

The "high" level of a group-factor is defined as the level in which all factors within that group are maintained at their high level. The "low" level of a group-factor is similarly defined. Although high and low level specification of a factor is arbitrary, the level anticipated to produce the larger response should be defined as the high level in order to reduce the possibility of group-factor effect cancellation.

Only the individual factors which comprise the significant first stage group-factors are carried over to the second stage to be tested individually using a Plackett-Burman design. Any factor which tests significant in the second stage is classified as active. Those factors not having a significant effect at the second stage and those factors not even carried over from the first stage are classified as inactive.

B. RANDOM BALANCE/PLACKETT-BURMAN STRATEGY

This strategy is also a two-stage strategy. The first stage consists of a K factor random balance design employing N runs, while the second stage

consists of a Plackett-Burman design follow-up.

In the random balance design each factor appears at its high and low levels N/2 times (N even). However, as remarked in Section II, no specific techniques for the analysis of random balance designs have been developed. Typically, each factor is considered separately, and some standard statistical analysis is applied to test for the presence of factor effects. Anscombe (1959) suggests using the ANOVA F-test, Welch's Randomization Test, or Tukey's Randomization Test. Each of these procedures admits to a relatively simple statistical analysis. More complex techniques could be devised at the possible expense of analysis simplicity, a consideration that might outweigh the advantages a complex procedure could offer. Exactly which procedure will result in the best performance of this strategy is an important issue worthy of research.

However, once a test of significance is decided upon, any factor testing significant in the random balance design is carried over to a Plackett-Burman design second-stage follow-up. Any factor which tests significant in this stage is classified as active. Those factors not having a significant effect at the second stage and those factors not even carried over from the first stage are classified as inactive factors.

C. DISCUSSION

The two strategies just discussed have been outlined in complete generality. To apply the group screening strategy in a particular situation, the group size, g, must be specified. Similarly, to apply the random balance/ Plackett-Burman strategy, the number of first-stage runs, N, must be specified. Thus, there exist many variations of these basic strategies depending on g

and N. To establish a notation denoting these variations, let GS_g denote the group screening strategy having a group size equal to g and let RB_c denote the random balance/Plackett-Burman strategy where $N = cK$, $0 < c < 1$.

For a given set of values for $\beta_1, \beta_2, \ldots, \beta_k$ and K in the factor screening model, the performance of the strategies RB_c and GS_g can be compared for various values of c and g. The notion of performance will, however, need to be quantified. One possible approach to devising a performance measure is presented in the next section.

IV. EVALUATING PERFORMANCE

In order to assess the effectiveness of a factor screening strategy or to compare competing strategies, some performance measure must be defined. There are, of course, certain characteristics of a screening strategy that any reasonable performance measure should take into account.

As Watson (1961) points out, the objectives of a screening strategy are:

(a) to detect as many of the active factors as possible,

(b) to declare active as few inactive factors as possible,

and (c) to achieve these aims with as few runs as possible.

Therefore, a good performance measure should reflect the extent to which the above objectives are satisfied. Given such a measure, strategies can then be objectively compared and evaluated.

As might be expected, there is no single screening strategy which will outperform every other strategy with respect to each of the above objectives. Indeed, simply consider the two strategies, both employing $R = 0$ runs, where in one strategy all factors are declared active and in the other strategy all factors are declared inactive. Neither strategy is uniformly better than the other. However, a strategy that is uniformly better than both of these trivial strategies would require $R = 0$ runs and no chance of factor misclassification!

Since objectives (a) and (b), which deal with factor classification, and objective (c), which deals with testing cost, are conflicting requirements, some balancing of objectives will be required in order to compare

competing strategies. In general, balancing the objectives will depend on:

(1) the relative importance of the objectives to the analyst,

(2) the tradeoffs in performance the analyst is willing to tolerate,

and (3) the consequences of violating any of the objectives.

Because the relative importance of these considerations will vary from situation to situation, so will a suitable performance measure. In short, in a given factor screening situation, criteria that meet the particular requirements of the situation will need to be specified by the analyst. In the following paragraphs, fairly general criteria will be developed to illustrate one possible approach to devising a suitable performance measure.

To begin, in factor screening two possible errors exist:

(i) declaring active an inactive factor,

and (ii) declaring inactive an active factor.

To measure the severity of these errors, a loss function can be used. In general, the selection of a suitable loss function is not trivial. However, for the moment, consider the class of functions

$$L(\underline{\beta}) = \sum_{i=1}^{K} w_i \delta_i / \sum_{i=1}^{K} w_i,$$

where

$$\delta_i = \begin{cases} 0 \text{ if the } i^{\underline{th}} \text{ factor is correctly identified} \\ 1 \text{ if the } i^{\underline{th}} \text{ factor is incorrectly identified,} \end{cases}$$

and w_i denotes the loss incurred ($w_i \geq 0$) if the $i^{\underline{th}}$ factor is misclassified.

Normally, the loss for misclassifying an active factor will be some

monotonic function of the true factor effect. For example, w_i might be defined as

$$w_i = \begin{cases} |\beta_i| \text{ if } |\beta_i| > \gamma \\ \gamma \text{ if } |\beta_i| \leq \gamma \end{cases}$$

for some γ (which may itself be a function of $\underline{\beta}$).

Since $L(\underline{\beta})$ is random for most screening strategies, expected loss, $E[L(\underline{\beta})]$, can be used as one basis for performance. Note that the loss function, $L(\underline{\beta})$, is normalized so that $0 \leq E[L(\underline{\beta})] \leq 1$. The quantity $(1 - E[L(\underline{\beta})])$ is therefore a measure of the efficiency of a screening strategy for classifying factors. Efficiency closer to one would indicate better performance on the average. However, as an overall standard of performance, classification efficiency, $(1 - E[L(\underline{\beta})])$, fails to account for the total number of runs, R, that a strategy requires.

Regarding this aspect of performance (objective (c)), let $\phi(R)$ represent the expense of employing R runs. Then the quantity

$$Q(\underline{\beta}) = E[\phi(R)]/\phi(K^*)$$

could be used as a measure of the relative testing cost of a strategy (relative to K*, the number of runs required for a Plackett-Burman design). Although a smaller testing cost would indicate better performance on the average, it is imperative that expected loss and relative testing cost be considered jointly in assessing the overall performance of a screening strategy (especially since $E[L(\underline{\beta})]$ and $Q(\underline{\beta})$ represent conflicting criteria). In some sense the problem is akin to the testing of a statistical hypothesis in which the probabilities of Type I and Type II error are both desired

small, but are inversely related.

As a result of these considerations, the analyst may want to specify the joint values of expected loss and relative testing cost that are acceptable. For any strategy within the restricted class satisfying the acceptability criteria, expected loss and relative testing cost might then be combined into some overall measure to further evaluate and compare performance. For example in comparing two acceptable strategies, the ratio $(1 - E[L(\underline{\beta})])/Q(\underline{\beta})$, or perhaps some linear combination of $L(\underline{\beta})$ and $Q(\underline{\beta})$ could be used.

V. REFERENCES

Anscombe, F. J. (1959). "Quick Analysis Methods for Random Balance Screening Experiments," Technometrics, Vol. 1, p. 195.

Booth, K. H. V., and Cox, D. R. (1962). "Some Systematic Supersaturated Designs," Technometrics, Vol. 4, p. 489.

Budne, T. A. (1959). "The Application of Random Balance Designs," Technometrics, Vol. 1, p. 139.

Hunter, J. S., and Naylor, T. H. (1970). "Experimental Designs for Computer Simulation Experiments, "Management Science, Vol. 16, p. 422.

Jacoby, J. E., and Harrison, S. (1962). "Multi-Variable Experimentation and Simulation Models," Nav. Res. Log. Quart., Vol. 9, p. 121.

Kleijnen, J. P. C. (1975a). "Screening Designs for Poly-Factor Experimentation," Technometrics, Vol. 18, p. 487.

Kleijnen, J. P. C. (1975b). Statistical Techniques in Simulation (Part II), Marcel Dekker, Inc., New York.

Li, C. H. (1962). "A Sequential Method for Screening Experimental Variables," Journal of the American Statistical Association, Vol. 57, p. 455.

Montgomery, D. C. (1979). "Methods for Factor Screening in Computer Simulation Experiments," Technical Report, Georgia Institute of Technology.

Patel, M. S. (1962). "Group Screening with More Than Two Stages," Technometrics, Vol. 4, p. 209.

Plackett, R. L., and Burman, J. P. (1946). "The Design of Optimum Multifactor Experiments," Biometrika, Vol. 33, 305.

Satterthwaite, F. E. (1959). "Random Balance Experimentation," Technometrics, Vol. 1, p. 111.

Smith, D. E. (1973). "Requirements of an 'Optimizer' for Computer Simulation," Nav. Res. Log. Quart., Vol. 20, p. 161.

Watson, G. S. (1961). "A Study of the Group Screening Method," Technometrics, Vol. 3, p. 371.

Youden, W. J., Kempthorne, O., Tukey, J. W., Box, G. E. P., and Hunter, J. S. (1959). "Discussion of the Papers of Messrs. Satterthwaite and Budne," Technometrics, Vol. 1, p. 157.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1 REPORT NUMBER 113-2	2 GOVT ACCESSION NO.	3 RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FACTOR SCREENING IN COMPUTER SIMULATIONS: CONSIDERATIONS IN PERFORMANCE EVALUATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dennis E. Smith and Carl A. Mauro		8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0650
9. PERFORMING ORGANIZATION NAME AND ADDRESS Desmatics, Inc. P. O. Box 618 State College, PA 16801		10 PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 274-317
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Arlington, VA 22217		12. REPORT DATE July 1980 13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution of this report is unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Factor Screening Computer Simulation Experimental Design		
20 ABSTRACT (Continue on reverse side if necessary and identify by block number) Although computer simulation is a major area in which factor screening situations are frequently encountered, adequate methodology has not been developed to resolve the factor screening problem. Within the constraint of a limited number of computer runs, a decision must be made about the selection of a screening strategy that will perform efficiently and effectively. This report reviews the major classes of screening designs (cont.)		

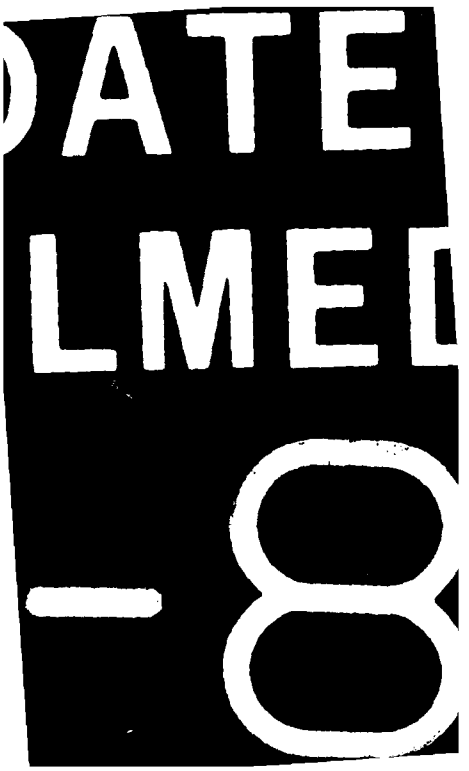
DD FORM 1 JAN 7 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(cont. 20.)

that have been suggested and recommends two strategies for further research. In addition, a fundamental screening model is defined, and a performance measure for evaluating screening strategies is developed.


DATE
LMED
-8